Current Research

Continuing Education Questionnaire, page 414
Meets Learning Need Codes 2000, 3000, 3010, and 9000

Methodology for Adding Glycemic Load Values to the National Cancer Institute Diet History Questionnaire Database

ANDREW FLOOD, PhD; AMY F. SUBAR, PhD; STEPHEN G. HULL, MS; THEA PALMER ZIMMERMAN, MS, RD; DAVID J. A. JENKINS, MD, PhD, DSc; ARTHUR SCHATZKIN, MD, DrPH

ABSTRACT
Background A growing interest exists in using glycemic index and glycemic load as potentially important exposures in investigations of risk for a variety of chronic diseases.
Objective We added values for glycemic index and glycemic load to the nutrient database of a commonly used dietary assessment instrument, the Diet History Questionnaire (DHQ).
Design The nutrient database for the DHQ is based on 4,200 individual foods reported by adults in the 1994-1996 US Department of Agriculture Continuing Survey of Food Intakes by Individuals (CSFII). This list was condensed into 225 nutritionally similar groupings of individual foods. Using published glycemic index values we assigned glycemic index values to each of the individual CSFII foods in these food groups. In cases where CSFII foods did not correspond tightly to foods with published glycemic index values, we used decision criteria to assign glycemic index values. We then calculated sex- and serving size-specific glycemic load for each of the 225 food groups using the weighted mean method. Quality assessments were made to help evaluate the success of this method for assigning glycemic load values.
Results Seventy-one percent of the top carbohydrate-contributing food groups had in excess of 90% of the CSFII mentions linked directly to a published glycemic index value (ie, no imputation was required), and 100% of these food groups had at least 50% of total mentions linked directly.
Conclusions Using this method, it is now possible to use DHQ responses to assess the associations between reported glycemic load and glycemic index and risk of many chronic diseases in epidemiologic studies.
J Am Diet Assoc. 2006;106:393-402.

A. Flood is an assistant professor, Division of Epidemiology and Community Health, University of Minnesota, Minneapolis. A. F. Subar is a research nutritionist, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD. S. G. Hull and T. P. Zimmerman are research nutritionists with Westat, Inc, Rockville, MD. D. J. A. Jenkins is a professor, Department of Nutritional Sciences, University of Toronto, Ontario, Canada. A. Schatzkin is chief of the Nutritional Epidemiology Branch and a senior investigator, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

Address correspondence to: Andrew Flood, PhD, Division of Epidemiology, University of Minnesota, 1300 S Second St, Suite 300, Minneapolis, MN 55454. E-mail: flood@epi.umn.edu

0002-8223/06/10603-0008$32.00/0
doi: 10.1016/j.jada.2005.12.008

The glycemic index, as developed in the early 1980s, was a tool to improve the management of glycemic control in patients with diabetes (1). The glycemic index is a measure of the glycemic effect of the carbohydrate in a particular food compared with an equivalent amount of carbohydrate in standard glucose or white bread (although the protein in foods also produces a glycemic effect, the acute response is much less than that from carbohydrate and for the purposes of overall dietary assessment, the focus of this article, it can be considered negligible). Determining the glycemic index of a food involves first feeding 50 g glucose to individual subjects and plotting the subsequent glycemic response during the following 2 hours. Next, subjects consume the test food in an amount that provides 50 g carbohydrate (for foods with low carbohydrate density, like vegetables, this can be problematic), and investigators plot the glycemic response to the test food. The glycemic index is the ratio ($\times$100) of the area under the curve for the glycemic response to the test food to the area under the curve for the glycemic response to glucose. To combine the qualitative and quantitative measures of carbohydrate, Liu and col-

Table 1. Example of a National Cancer Institute (NCI) Diet History Questionnaire (DHQ) line item with links to multiple Continuing Survey of Food Intakes by Individuals (CSFII) foods but with only one item in the Glycemic Index Table (16) linked to all CSFII foods

CSFII food group on NCI DHQ	Foods from CSFII in the NCI DHQ database	No. of times mentioned by respondents to CSFII
Oranges, tangerines, tangelos	Orange, raw	941
	Orange, mandarin, canned or frozen, not specified as to sweetened or unsweetened; sweetened, not specified as to type of sweetener	12
	Orange, mandarin, canned or frozen, in light syrup	2
	Orange, mandarin, canned or frozen, drained	9
	Tangelo, raw	10
	Tangerine, raw	81

leagues and Salmeron and colleagues (2-4) developed the concept of glycemic load. The glycemic load of a serving of a specific food is simply the product of its glycemic index (÷100) and the grams of carbohydrate from a single serving of that food. Glycemic load can be thought of as an indicator for the gram equivalents of pure glucose, in terms of glycemic response, delivered by a single serving of a specific food. The glycemic load of the diet is simply the sum of the glycemic loads for the total servings of all carbohydrate-containing foods consumed on average per day.

A growing number of epidemiologic studies have investigated glycemic index and glycemic load as potentially important exposures in investigations of risk for a variety of chronic diseases, including diabetes (3,4), cancer (1,5-10), and cardiovascular disease (2), as well as for intermediate markers of risk such as serum lipids (11-14), glycated hemoglobin (12), and high-sensitivity C-reactive protein (15).

Despite this growing interest in glycemic index and glycemic load as markers of risk factors for disease, the methods for assessing these exposures in an epidemiologic context are neither well established nor consistently applied across studies. Epidemiologic studies of nutritional exposures often rely on food frequency questionnaires (FFQs) to estimate usual dietary intake and use algorithms to convert reported frequency and portion size from these FFQs into estimates of nutrient intake. These algorithms assign nutrient values based on databases developed from national food composition tables, but a major complication in studying glycemic index and glycemic load is the lack of any national food composition table that includes these variables. Thus, it is necessary for investigators wishing to study glycemic load to add glycemic load values to their FFQ databases without benefit of standard values compiled in a single food composition table. We added glycemic load values to the National Cancer Institute (NCI) Diet History Questionnaire (DHQ) database, and in an effort to shine light on the precise methods used for assigning these values and to discuss the complications and limitations associated with doing so, we provide here a detailed description of the method. It is our hope that providing this information, information that can be adapted for use with other FFQs, will contribute to a standardization of methods for adding glycemic load values to FFQ databases and thus remove some of the mystery surrounding how investigators arrive at estimates of dietary glycemic index and average daily glycemic load for subjects in large epidemiologic studies.

METHODS

The NCI DHQ

The DHQ is a widely used FFQ developed by investigators at NCI. Its food list, portion sizes, and nutrient database are based on responses from 10,019 adults aged 19 years and older who completed one or two 24-hour dietary recalls administered as part of the US Department of Agriculture's (USDA's) 1994-1996 Continuing Survey of Food Intakes by Individuals (CSFII). On their dietary recalls, these respondents listed a total of 5,261 distinct food items that NCI investigators subsequently categorized into 336 cognitively and nutritionally similar, mutually exclusive groups. These food groups contained as few as one of the individual food items mentioned on 24-hour recalls by respondents to the CSFII to composites of as many as 161 foods. As one example (Table 1), the food group "oranges, tangerines, and tangelos" represents a composite of six different foods mentioned at least once in the CSFII (orange, raw; tangerine, raw [including mandarin orange, Satsuma]; oranges, mandarin, canned or frozen, sweetener not specified; tangelo, raw; oranges, mandarin, canned or frozen, drained; oranges, mandarin, canned or frozen, in light syrup). After evaluating these groupings in terms of overall nutrient contribution and cognitive clarity, 225 of these CSFII food groups were identified. We refer to these hereafter as DHQ food groups. The 225 DHQ food groups form the basis of the line items in the DHQ instrument.

Compilation of Published Glycemic Index Values

As mentioned, no national food composition table contains values for glycemic index. Lack of such a table necessitates the use of alternative sources for assigning glycemic index values to food items in the DHQ database. In 2002, Foster-Powell and colleagues (16) compiled a list of all glycemic index values published between 1981 and 2001. Also in their compilation, Foster-Powell and colleagues included unpublished data from Sydney University's Glycemic Index Research Service and other labs that were verified to meet strict standards of methodological rigor. The resulting list (hereinafter referred to as the Glycemic Index Table) included more than 1,300 separate published (or quality-confirmed, unpublished) glycemic index values corresponding to more than 750 individual food items. The foods were organized into 22 food groups: Bakery Products, Beverages, Breads, Breakfast Cereals and Related Products, Breakfast Cereal Bars, Cereal Grains, Cookies, Crackers, Dairy Products and Alternatives, Fruit and Fruit Products, Infant Formula and Weaning Foods, Legumes and Nuts, Meal-Replacement Products, Mixed Meals and Convenience Foods, Nutritional-Support Products, Pasta and Noodles, Snack Foods and Confectionery, Sports Bars, Soups, Sugars and Sugar Alcohols, Vegetables, and Indigenous or Traditional Foods of Different Ethnic Groups. There were no glycemic index values in the table for meat, poultry, fish, avocados, salad vegetables, cheese, or eggs because these foods contain little or no carbohydrate, making it very difficult for a person to consume a serving of any of them that contained 25 to 50 g available carbohydrate (as would be required for a clinical determination of glycemic index). This Glycemic Index Table was the source for the glycemic index values we included in the DHQ database.

Linking Published Glycemic Index Values to DHQ Food Groups

The first step in creating glycemic load values for inclusion in the DHQ database was the linkage by a nutritionist of individual foods in the Glycemic Index Table to each of the 4,220 individual CSFII foods that corresponded to the 225 food groups comprising the line items in the DHQ. The method of linkage was by manual review of the Glycemic Index Table to identify those foods that, in the judgment of the investigators, were the best matches for each of the CSFII foods. As this description suggests, the criteria we used in linking foods from the Glycemic Index Table to CSFII foods were necessarily subjective. Unfortunately, there simply are no objective criteria for saying that food X is the best match for food Y with respect to glycemic index. A nutrient comparison among foods would add little because it is not the amount of the carbohydrate in a food that determines its glycemic index, but rather the quality of that carbohydrate in terms of how quickly it is absorbed. To confirm the validity of these linkages, a second dietetics professional reviewed these initial matches to determine their appropriateness and to determine if further matches could be made. To allow users to review the final linkages we made, the DHQ Web site (http://riskfactor.cancer.gov/tools/glycemic/) includes a compilation of all the links from the Glycemic Index Table published by Foster-Powell and colleagues (16) to the CSFII foods.

In the case of foods with multiple entries in the Glycemic Index Table (ie, where multiple published values exist for a single food), we used the mean value of the listed foods. For example, the Glycemic Index Table has six entries listed under the heading "oranges, raw," and we simply used the mean from each of these independently published values (42.0) when linking "oranges, raw" to the CSFII foods. For foods with multiple entries in the Glycemic Index Table but for which one or more of the listed foods (or its equivalent) was not typically consumed in the United States, the non-American foods were excluded from the mean glycemic index value.

The linkage process could in principle be completely straightforward with each of the CSFII foods linked directly to a single, unique, and perfectly comparable food in the Glycemic Index Table. For some CSFII foods, it was in fact very easy to find a Glycemic Index Table food that matched (eg, "Milk, cow's, fluid, whole" in the CSFII was matched to "Milk, full-fat" in the Glycemic Index Table). In practice, however, there were many examples where this was not the case. In situations where the links were not as clean as this idealized scenario, we were forced to make judgments about the best linkages possible given the finite and not necessarily compatible list of foods in the Glycemic Index Table. To handle the more difficult cases we employed an algorithm for assigning glycemic index values (Figure). It is useful to note that we did not proceed through all nine steps in the algorithm for a single CSFII food and then move on to the next food. Instead, we proceeded through Step 1 for all 4,220 CSFII foods, and for those without a link after Step 1, we moved on to Step 2, and for those without a link after Step 2 we proceeded to Step 3, and so on. A description of the various steps in this algorithm appears below:

Step 1. Determine if the CSFII food has a direct link to a food in the Glycemic Index Table. If yes, assign the glycemic index value for the food in the Glycemic Index Table. If not:

Step 2. Determine if there is a closely related food in the Glycemic Index Table. Foods that lacked a direct link to a food in the Glycemic Index Table nonetheless were often nutritionally similar to foods that did have a listing in the Glycemic Index Table. If, in the judgment of the dietetics professionals making the glycemic index assignments, the similarity was sufficiently close, we imputed the glycemic index value of the related food in the Glycemic Index Table for the glycemic index value of the CSFII food in question. If there was no closely related food in the Glycemic Index Table:

Step 3. Determine if the CSFII food is a vegetable. Vegetables typically have very low carbohydrate density, making a clinical determination of their glycemic index problematic. In practical terms, the low carbohydrate density makes it difficult for test subjects to consume enough of most vegetables to account for 50 g carbohydrate. Thus, lacking specific, clinically determined glycemic index values for most vegetables, we calculated a simple mean of the glycemic index values for vegetables that were listed in the Glycemic Index Table and used this as an imputed value for all other vegetables. Although this is a crude solution, the grams of carbohydrate per serving for these vegetables is very low, meaning that their contribution to glycemic load (see below) is small, and therefore the error introduced by this imputation will have relatively minor consequences. (It is important to note that high carbohydrate vegetables such as potatoes generally have well-established glycemic index values and have listings in the Glycemic Index Table. We did not include these starchy vegetables in the definition of "vegetable" as applied in this step.) If the CSFII food was not a vegetable:

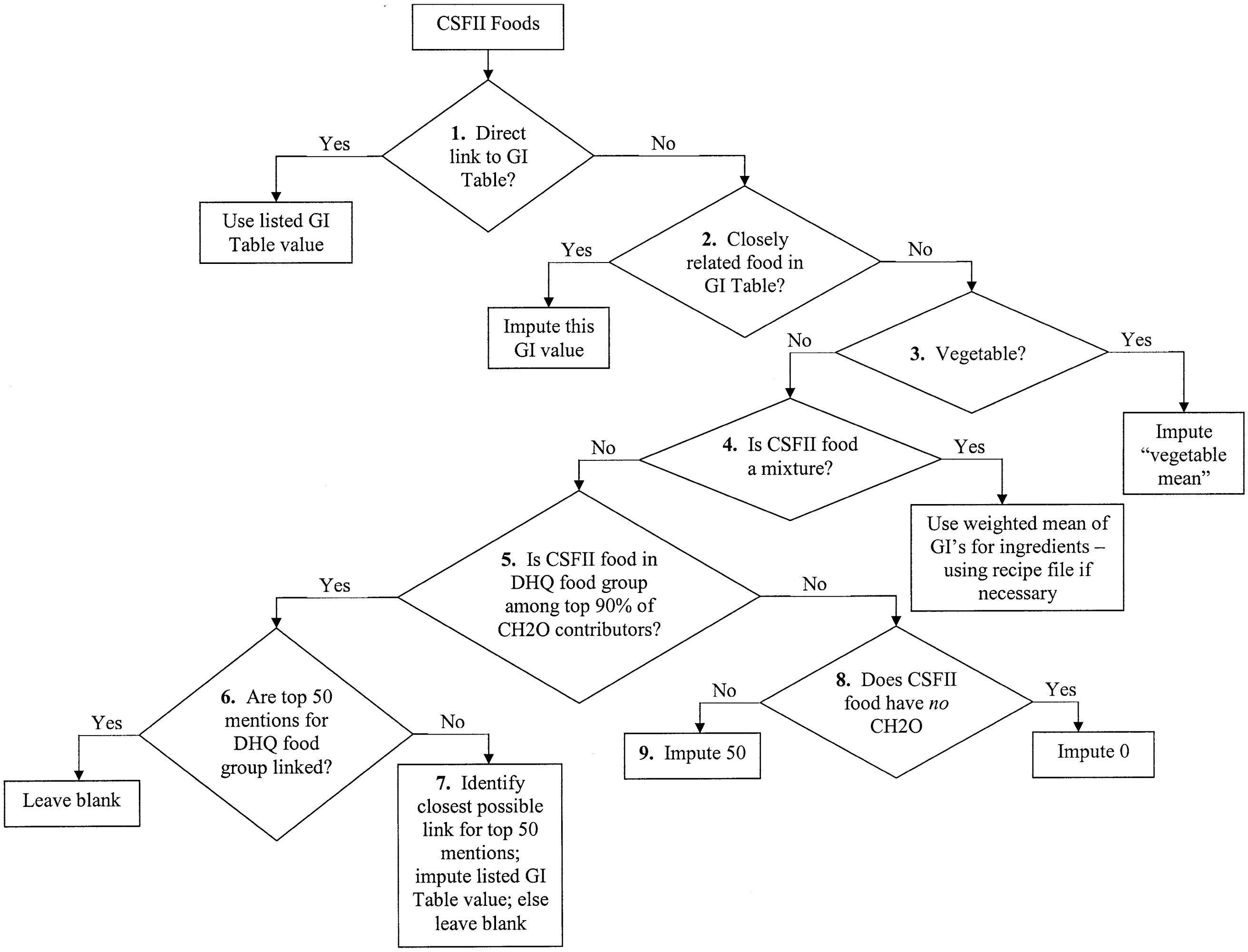

Figure. Nine-step Glycemic Index (GI) Table–Continuing Survey of Food Intakes by Individuals (CSFII) linkage algorithm used to assign GI values to foods. DHQ=Diet History Questionnaire. CH2O=Carbohydrate.

Step 4. Determine if the CSFII food is a simple mixture of foods in the Glycemic Index Table or if the CSFII food is a mixed dish in the CSFII recipe file. In the case of a simple mixture, we matched each component of the CSFII food to the appropriate Glycemic Index Table food. For example, "Peas and carrots, cooked, NS as to form, fat not added in cooking" in the CSFII was matched to both "Green peas" and "Carrots" in the Glycemic Index Table. We calculated the glycemic index for this CSFII food to be simply the mean of the glycemic indexes for "Green peas" and "Carrots." For more complex mixed dishes that had listings in the CSFII recipe file, we calculated a glycemic index value using a mean of the glycemic index values for the ingredients in that recipe weighted by their contribution to the total carbohydrate in the mixed dish (Table 2). If the CSFII food was not a mixed dish:

Step 5. Determine if the CSFII food is in a DHQ food group that is in the top 90% of carbohydrate-contributing food groups to the diet as measured by the CSFII (see Table 3). For CSFII foods in DHQ food groups making a minor contribution to the carbohydrate in the diet, we allowed for a less stringent method of glycemic index assignment (see Step 8), but for the major carbohydrate contributors we made extra effort to assign a more solid glycemic index value:

Step 6. Determine if, among the foods in the DHQ food group containing the food we are trying to link, the most frequently mentioned foods, through at least 50 mentions, have links to the Glycemic Index Table. A mention is defined as an eating occasion reported on the CSFII 24-hour recalls (in the example above for the DHQ food group "oranges, tangerines, and tangelos," there were 941 mentions in the CSFII 24-hour recalls of "oranges, raw"; 81 of "tangerine, raw (include mandarin orange, Satsuma)"; and 12 of "oranges, mandarin, canned or frozen, sweetener not specified"). If the top 50 mentions in the DHQ food group have a solid link to the Glycemic Index Table, we left the glycemic index value blank for foods making up a small proportion of the mentioned occurrences of consumption for CSFII foods in that DHQ food group. For example, in the DHQ food group "RTE (ready-to-eat) cereal, highly fortified," the least frequently men-

Table 2. Calculating the glycemic index for a mixed dish in the Continuing Survey of Food Intakes by Individuals (CSFII) recipe file

CSFII mixed dish	Foods from CSFII in the recipe file for this mixed dish	Grams of carbohydrate contributed by recipe food to a single serving of the mixed dish	Links to recipe food from Glycemic Index Table (with linked glycemic index value in bold)
Beef stew with potatoes and vegetables (including carrots, broccoli, and/or dark-green leafy), tomato-based sauce	Beef, chuck, arm pot roast, choice, separable lean only, raw	0	Not applicable
	Carrots, raw	14.6	Carrots; mean of 4 studies: **47**
	Onions, raw	9.5	Vegetable, no food specified; imputed mean: **32**
	Peas, green, canned, regular pack, drained solids	14.3	Green peas; mean of 3 studies: **48**
	Tomatoes, red, ripe, canned, whole, regular pack	4.6	Tomato juice, canned, no added sugar (Berri Ltd, Berri, Australia): **38**
	Potatoes, raw, flesh	43.9	Potato, boiled, no food specified; imputed mean: **72**
	Water, municipal	0	Not available
	Salt, table	0	Not available
	Wheat flour, white, all-purpose, enriched, bleached	17.9	Baguette, white, plain (France): **95**
Calculated glycemic index (weighted mean) for mixed dish: 64.1			

tioned CSFII food, "King Vitaman," (Quaker Oats Co, Chicago, IL) accounted for only two of the total 209 mentions for that food group, whereas the most frequently mentioned food, "Total," (General Mills, Inc, Minneapolis, MN) accounted for 146 of those mentions (easily covering the top 50 mentions all by itself). Lacking a Glycemic Index Table listing for King Vitaman, the glycemic index value for this CSFII food was left blank. Foods with a blank value for glycemic index were excluded from the glycemic load calculation for that DHQ food group (see below for a description of glycemic load calculation methods). If foods that contribute to the top 50 mentions for the DHQ food group were not linked, then:

Step 7. Identify the closest possible link in the Glycemic Index Table. Despite lacking a closely related food (as might have been identified in Step 2 above), there were nonetheless foods that had at least some nutritional similarity to the CSFII food we were trying to link. In these cases, we assigned the closest possible food in the Glycemic Index Table. After linking all CSFII foods without a closely related food in the Glycemic Index Table using this method, we imputed these imperfect glycemic index links until we had accounted for the top 50% of the CSFII mentions for that food group.

Step 8. For CSFII foods that are not closely linked to a food (or foods, in the case of mixed dishes) in the Glycemic Index Table and that are not in a DHQ food group in the top 90% of carbohydrate contributors, determine if its DHQ food group has 0 g carbohydrate. For foods lacking carbohydrate, impute a glycemic index value of zero. For those that do have carbohydrate:

Step 9. Impute a glycemic index value of 50. Rather than undertake a more elaborate imputation process for these minor contributors to dietary carbohydrate, we simply imputed a medium glycemic index value of 50.

Calculating Glycemic Load and Dietary Glycemic Index

Having linked glycemic index values to as many CSFII foods as possible, we were then able to calculate glycemic load per serving by portion size and sex for each line item in the DHQ. We began by calculating a glycemic load value for every mention of a CSFII food in the DHQ. This calculation was simply the product of the available carbohydrate content (in grams) for each individual mention of that food on any respondent's 24-hour recall reported in the CSFII and the glycemic index (÷100) for the food consumed on that eating occasion (excluding eating occasions that involved foods for which we had a missing value for glycemic index). This is equivalent to calculating a nutrient value for intake of any food reported on a 24-hour recall. Next, each mention was classified as being of a small, medium, or large portion size based on cutpoints determined from previous work for the DHQ database (17). These cutpoints defined medium portion sizes as between approximately the 25th and 75th per-

Table 3. Top 90% of carbohydrate contributors among Diet History Questionnaire (DHQ) food groups with listing of carbohydrate content and calculated glycemic load for a man's medium serving of each food

DHQ food group	% of total carbohydrate	Cumulative % of total carbohydrate	% of CSFII mentions linked[a]	Grams of carbohydrate per man's medium serving	Glycemic load per man's medium serving
Soft drinks, regular, caffeinated	8.96	8.96	91	40.0	23.1
Breads/rolls, white	7.37	16.32	99	23.5	15.8
Fruit drinks, regular	3.15	19.47	74	39.6	24.7
Bread/rolls whole grain	2.95	22.43	100	23.8	13.5
Orange/grapefruit juice, all	2.57	25.00	100	23.5	11.5
Potatoes, fried	2.56	27.56	97	33.4	23.0
Soft drinks, regular, decaf	2.50	30.05	71	38.6	24.2
Ready-to-eat cereal, good fiber	2.37	32.42	88	36.1	20.7
Ready-to-eat cereal, other	2.31	34.73	83	38.9	28.2
Cakes, regular	2.31	37.04	99	42.3	18.8
Rice/grains, no fat added	2.11	39.15	100	45.6	28.1
Cookies, brownies	1.98	41.12	83	22.4	11.6
Bananas	1.85	42.98	100	27.2	12.7
Pasta, no fat added	1.67	44.65	98	79.8	32.7
Potato/corn/other chips	1.66	46.31	100	18.0	9.5
Donuts, sweet rolls, Danish, pop tarts	1.65	47.96	100	34.5	22.6
Pizza, with meat	1.54	49.50	100	48.4	14.9
Ice cream, regular	1.52	51.02	83	31.1	18.3
Beer	1.38	52.40	100	18.2	6.3
Sugars/honey, all in coffee or tea	1.36	53.76	100	7.0	4.7
Crackers	1.31	55.07	96	16.6	11.0
English muffin/bagel	1.30	56.37	100	28.9	19.7
Candy, chocolate	1.16	57.53	69	23.8	9.7
Potatoes, white, no fat added	1.16	58.69	100	26.5	17.5
Other juice	1.08	59.77	91	31.1	13.1
Beans, fat added	1.02	60.79	96	27.7	7.1
Mexican mixtures, all	1.01	61.80	53	34.3	12.8
Chicken, mixtures	0.99	62.79	60	24.1	14.2
Apples	0.93	63.72	100	20.1	6.3
Macaroni and cheese	0.92	64.64	88	51.6	31.4
Pancake, waffle, French toast	0.90	65.53	100	36.4	24.0
Hot breakfast cereals, no fat added	0.88	66.42	100	30.8	16.3
Rice/grains, fat added	0.88	67.30	92	38.6	23.3
Milk, 2% not in coffee/tea	0.82	68.12	100	14.4	4.3
Beef stews/pot pies/mixtures	0.79	68.91	54	35.4	15.9
Biscuits, all	0.74	69.65	95	23.4	20.9
Cornbread/muffins	0.74	70.38	100	31.3	22.1
Pasta, meat/fish sauce	0.73	71.12	93	43.8	20.9
Pies, fruit	0.72	71.84	100	49.1	27.8
Maple syrup on pancakes, etc.	0.72	72.56	72	37.9	7.0
Pretzels, all	0.69	73.25	100	25.9	20.6
Pasta, fat added	0.66	73.91	99	68.9	27.6
Milk, whole not in coffee/tea	0.64	74.55	100	14.2	3.9
Milk, nonfat/skim not in coffee/tea	0.62	75.17	100	16.0	5.3
Candy, not chocolate	0.62	75.79	62	21.0	16.1
Jams, jelly, regular	0.60	76.39	100	10.2	5.1
Pizza, without meat	0.60	76.98	96	47.5	25.7
Soups, with vegetables	0.59	77.57	89	20.9	7.5
Lasagna, ravioli, shells, and so on	0.58	78.16	73	51.7	19.3
Popcorn	0.58	78.73	91	22.0	12.5
Frozen yogurt, ices, sorbet, and so on	0.57	79.30	89	26.4	14.8
Muffins/dessert breads, regular	0.56	79.87	100	30.2	16.5
Coffee, regular, no cream/sugar	0.54	80.40	100	2.0	1.0
Pasta, meatless red sauce	0.52	80.92	100	43.7	15.9

(continued)

Table 3. Top 90% of carbohydrate contributors among Diet History Questionnaire (DHQ) food groups with listing of carbohydrate content and calculated glycemic load for a man's medium serving of each food (continued)

DHQ food group	% of total carbohydrate	Cumulative % of total carbohydrate	% of CSFII mentions linked[a]	Grams of carbohydrate per man's medium serving	Glycemic load per man's medium serving
Beans, no fat added	0.51	81.44	99	27.2	6.9
Ice cream/ice milk, low fat	0.50	81.94	71	35.2	15.6
Yogurt, all	0.49	82.42	100	29.6	8.6
Corn, no fat added	0.49	82.91	100	23.8	10.3
Soups, broth with noodles/rice	0.49	83.40	86	15.9	6.7
Sugars/honey, all not in coffee/tea	0.47	83.87	100	6.5	4.4
Cookies, brownies, low fat	0.43	84.29	96	26.1	14.3
Chili	0.39	84.68	100	28.0	7.0
Pies, cream/custard/other	0.39	85.07	100	49.8	28.4
Oranges, tangelo, and so on	0.39	85.45	100	14.6	4.9
Peaches/nectarines/plums	0.38	85.83	100	13.9	5.4
Fruit salads/other fruits	0.37	86.20	82	12.5	5.7
Misc. syrups, toppings	0.36	86.56	79	34.3	7.8
Puddings/custards	0.35	86.91	100	38.5	17.6
Milkshakes/sodas	0.34	87.25	94	73.6	33.5
Potato salads	0.34	87.58	100	30.7	18.3
Beef, ground, meatballs/loaf/mixtures	0.33	87.91	55	13.4	4.8
Milk, 2% in cereal	0.33	88.24	100	9.2	2.7
Stuffing/dumplings, all	0.30	88.54	78	28.0	20.4
Tomato catsup	0.30	88.83	100	4.7	1.7
Tomatoes, raw	0.29	89.13	100	2.1	0.6
Other melon	0.29	89.42	99	14.1	9.2
Fish fried, fat added	0.28	89.70	100	12.9	11.9
Grapes, all	0.27	89.97	100	18.4	8.0
Pasta salad	0.26	90.23	63	33.8	14.9

[a]In all cases, for DHQ food groups in the top 90% of carbohydrate contributors, 100% of the top 50 mentions were linked. CSFII=Continuing Survey of Food Intakes by Individuals.

centile of total gram weight intakes (for men and women combined) for all foods contained within a DHQ food group. We then analyzed, by sex, all mentions making up each DHQ food group to assign glycemic load values to each of six categories: man small, man medium, man large, woman small, woman medium, and woman large. This was done by calculating the mean of the glycemic loads for the individual mentions contained within them. These glycemic load values can be used in the DHQ database to calculate overall daily glycemic load based on DHQ reported frequency and portion size by sex across all items on the questionnaire.

It is important to note that because DHQ food groups are typically made up of multiple CSFII foods, and because the people who eat large portion sizes may eat a different mix of those CSFII foods compared with those who eat small portion sizes, and because women may eat a different mix from men, no two sex- and portion size-specific categories within any DHQ food group will necessarily contain the same relative amounts of the different CSFII foods.

An example of how this worked for a single food group follows: As described in Table 1, CSFII respondents reported on the 24-hour recalls 1,055 eating occasions in which they consumed a food in the DHQ food group "oranges, tangerines, tangelos." These 1,055 mentions of an eating occasion that included one of these foods are distributed among the six CSFII foods that comprise this DHQ food group. We calculated the glycemic load for each mention by multiplying the grams of carbohydrate for that serving-size-specific mention and the glycemic index for the CSFII food. For example, one of the 941 mentions of "oranges, raw" was from a man who reported eating a medium-sized orange. Using USDA data, a medium orange provides 12.29 g available carbohydrate, so multiplying this value by 0.42 (the glycemic index for "oranges, raw" divided by 100) we arrive at a value of 5.16 g for the glycemic load of this mention. We repeated this process for all of the remaining 1,054 mentions in this DHQ food group. Serving sizes had been previously determined for the DHQ food group "oranges, tangerines, tangelos." For men, mentions with gram weights above 135 g (72nd percentile) were defined as large, and mentions with gram weights below 95 g (20th percentile) were defined as small. In the case of the "oranges, tangerines, tangelos" food group, the cut points were the same for women and men. For each sex-specific serving size category, we determined the mean glycemic load among all mentions in that category. For men, the mean of the glycemic load values for the 71 mentions of CSFII foods in the "oranges, tangerines, tangelos" food group in the small serving size category was 2.47. For the medium category it was 4.89, and for the large category it was 8.50. Among women, the mean glycemic load values for the mentions in the small, medium, and large categories were 2.43, 4.80, and 8.00, respectively. These values are the sex and serving-size specific glycemic load values that enter the final DHQ database.

Although we do not list in this article each of the sex- and serving size-specific glycemic load values for each of the DHQ food groups, these are posted on the DHQ Web site (http://riskfactor.cancer.gov/tools/glycemic/).

Methodological Note on Accounting for Dietary Fiber when Using the USDA Nutrient Databases. In the USDA food composition tables used to compute nutrient values for CSFII, the carbohydrate value includes both available (ie, digestible) carbohydrate and dietary fiber. Because glycemic load is meant as an indicator of the glycemic effect of food, and glycemic effect is inherently a function of the carbohydrate available for digestion and absorption, for the purposes of our glycemic load calculations, we defined carbohydrate to be the USDA-based value for grams of carbohydrate per serving minus the USDA value for grams of dietary fiber per serving. Strictly speaking, available carbohydrate excludes not just dietary fiber but also resistant starch, but the USDA tables include most resistant starch in their definition of fiber, so subtracting the USDA-based fiber value from total carbohydrate is a reasonable approach. Failure to remove fiber from the carbohydrate value used in these calculations would result in overestimation of the glycemic load from any food containing fiber or resistant starch.

Quality Assessment

We used several diagnostic measures to assess the quality of the methods described above for adding glycemic load values to the DHQ nutrient database. To judge the quality of the linkage process that assigned glycemic index values to each of the CSFII foods, first we determined the proportion of CSFII foods that we linked at each stage in the 10-step process described above. Although it is not possible to quantify, the glycemic index link procedures certainly have varying levels of quality associated with the strength of the links we made at each step. In relative terms, it is possible to say that Step 1 (direct link to a food in the Glycemic Index Table) clearly had the highest quality, Step 10 (arbitrary value of 50 imputed for foods with no links to the Glycemic Index Table and few mentions in a DHQ food group with low contribution to total carbohydrate) was the least solid, and the other steps fell in between. Next we determined what proportion of the CSFII mentions in each DHQ food group had links to foods listed in the Glycemic Index Table. Finally, we determined, for each DHQ food group, what percent of the top 50 CSFII mentions had links to foods in the Glycemic Index Table. For these analyses we focused on the top 90% of carbohydrate contributors among the DHQ food groups based on cumulative sum of carbohydrate contributed by each food group.

RESULTS

In assessing the quality of the linkages from the DHQ food list to the foods contained in the table of published glycemic index values, we determined that the method described above resulted in 40.2% of the total CSFII mentions (from 32.0% of CSFII foods) being linked directly to a Glycemic Index Table food (ie, they were linked at Step 1 in the linkage algorithm). Continuing through the algorithm, 19.8% of the mentions were linked at Step 2 (a closely related food in the Glycemic Index Table), 7.1% at Step 3 (a vegetable), 4.4% at Step 4 (a food mixture), 22.9% at Step 7 (the closest link possible for CSFII foods among the top 50 mentions in a DHQ food group contributing in the top 90% of carbohydrate), and 5.6% at Step 8 (a food that has no carbohydrate).

Despite this growing interest in glycemic index and glycemic load as markers of risk factors for disease, the methods for assessing these exposures in an epidemiologic context are neither well established nor consistently applied across studies.

When we concentrated on the DHQ food groups contributing the top 90% of carbohydrate in the DHQ, the proportion of CSFII foods with linkages at Step 1 in the linkage algorithm was even higher, with 62.1% of these mentions linked directly to a food in the Glycemic Index Table. When looking from the perspective of single DHQ food groups (rather than from that of all the CSFII foods in all food groups), 9.0% of the DHQ foods had 100% of their mentions linked directly (ie, at Step 1). More significantly, though, nearly half of the DHQ food groups in the top 90% of carbohydrate contributors had 100% of their mentions linked (ie, did not rely on imputed values), and 71% of the top carbohydrate contributors had in excess of 90% of their mentions linked. Only 22.9% of the total mentions for these major contributors to total carbohydrate were linked at Step 6 or 7 (ie, at steps where the quality was comparatively low). Finally, with the exception of beer (for which we used a single imputed value), 100% of the DHQ food groups contributing the top 90% of carbohydrate had at least 50% of their total mentions linked.

To focus on the most frequently mentioned foods within these top carbohydrate contributors, we ranked the CSFII foods with each DHQ food group based on the number of mentions each contributed to the total number of mentions for that group. In this way we determined that 100% of the DHQ food groups that contributed the top 90% of carbohydrate had links for 100% of their top 50 mentions. This means that the most frequently consumed CSFII foods within any DHQ food group that was a significant contributor of carbohydrate always had a high-quality link to the Glycemic Index Table.

DISCUSSION

We have described a method for adding glycemic load values to the nutrient database for a widely used FFQ, the

DHQ. Despite the lack of a national nutrient database with glycemic index values, we were able to add glycemic load to the DHQ database, and quality indicators suggest that we were successful in assigning high quality glycemic load values for almost all DHQ line items with significant carbohydrate contribution to the diet.

Of course, we were not 100% successful in linking the highest quality values to all the CSFII foods that form the basis of the DHQ nutrient database. Because it was necessary to rely on a reference table containing 750 foods with clinically determined glycemic index values, it was simply not possible to link all 4,220 CSFII foods directly. Nonetheless, for food groups that contributed significantly to the carbohydrate in the DHQ, and for the foods that were most frequently consumed in these food groups, we were able to assign high quality linkages to the published glycemic index values.

This process of adding glycemic load values to the DHQ nutrient database was obviously not without error. The error could enter the database glycemic load values we assigned to DHQ food groups at three stages. The first and most relevant to the issues covered in this article is the error in the linkages we made from the Glycemic Index Table to the CSFII foods. The nature of the linkages from the Glycemic Index Table to the CSFII foods, however, is qualitatively different from the linkages of nutrients in the USDA database. As described, with only 750 foods in the Glycemic Index Table and 4,220 foods in the CSFII food list that applies to the DHQ food groups, some gaps were inevitable. Nonetheless, the quality assessments we performed indicate that the overall strength of the linkages is quite high. Most of the gaps were for foods with little or no carbohydrate and would thus have minimal influence on any dietary analyses of glycemic load that used this database. For those foods contributing the top 90% of the carbohydrate in the DHQ, almost all had the great majority of their mentions linked to the Glycemic Index Table, and all of the most frequently consumed foods covered by all of these DHQ food groups that were major carbohydrate contributors had a link to the Glycemic Index Table. Although there still remain some gaps, and the closing of these gaps would naturally improve the quality of the glycemic load values in this database, the linkage methods we used were of sufficient quality to give confidence to those who would use the DHQ as a dietary assessment instrument in analyses of glycemic load.

The second source of error is in the overall calculation of the glycemic load values for line items on the DHQ. The error we contend with in the actual calculation of the sex– and portion size–specific glycemic load values would be no different from the error involved in the assignment of any nutrient value to the DHQ food groups using nutrient values (from the USDA database, for example) linked to the CSFII foods. To the extent that the method of assigning sex– and portion size–specific nutrient values is limited, so is the method of assigning glycemic load values. Yet this method has performed well for other nutrients and has been shown to be superior to possible alternative approaches (17).

A third way in which error could enter the database values for glycemic load would be in the clinical determination of glycemic index values for the 750 foods listed in the Glycemic Index Table. All of the linkages we made and the glycemic load values we calculated naturally depend on the glycemic index numbers presented in that table. Foster-Powell and colleagues (16) discussed some of the limitations associated with the laboratory methods used to arrive at glycemic index values for specific foods in the introduction to their 2002 update of the international table of glycemic index values, and others have raised additional issues (18). Important considerations related to laboratory methods include the use of arterial blood rather than venous blood for measuring glucose response, the standardization of diet and lifestyle practices in subjects in the days leading up to the clinical test, the use of diabetic vs healthy subjects (many of the early studies used diabetic subjects because the concept of glycemic index was originally intended for use in management of diabetes), and the use of white bread, a food that may be variable itself in terms of glycemic index and available carbohydrate per serving (although there is no specific evidence documenting this variability), instead of glucose as the standard food in glycemic index tests. Clearly, there are many complications of laboratory methodology that impinge on the reliability of the glycemic index values we have available to us in the Glycemic Index Table. But these (or comparable) issues concerning the reliability of laboratory methodology are common to the measurement of all nutrient values that dietetics professionals use in constructing their databases. Even though nutrients are measured in foods rather than in terms of a specific biological response to a food, there will always be error associated with those measurements. We may observe greater variability in the glycemic index values from test to test than we do for nutrients, but increased standardization and continued refinement of methods will help address this question. And mere variability in glycemic index values should not in and of itself be an indication that they are somehow qualitatively different from standard nutrient values commonly used in constructing nutrient databases, because there is considerable variability in measured values for nutrients as well (19,20). Just as there are continuing efforts to enhance the quality of nutrient measurements for use in nutrient databases, so will there be for glycemic index measurements, and as the measurements are refined, databases will be improved.

CONCLUSIONS

The methods described here represent an initial attempt to add glycemic load values to the nutrient database for a widely used FFQ, the DHQ (for more information about use of this instrument, see the DHQ Web site: http://riskfactor.cancer.gov/DHQ/). In the process of developing the methods for doing so, we established a series of assumptions and created decision rules to direct the final calculation of sex– and portion size–specific glycemic load values for the 225 DHQ food groups. To make clear how we arrived at these values, we have published our methods in detail. With time and with consultation from others working in the field, it will inevitably pass that we will adjust and refine these methods to improve the quality and utility of the glycemic load values in the database. Furthermore, glycemic load databases will improve as the tables of published glycemic index values are updated with new data using improved laboratory methodologies for an even broader range of individual foods. Nonetheless, the assessments we performed to evaluate the linkages of Glycemic Index Table foods to DHQ food groups strongly suggest that the glycemic load values we assigned for this nutrient database are of good quality, and as such, our method will make it possible to use the DHQ in large epidemiologic studies of the glycemic effects of food on many chronic disease outcomes. We developed our methods to be generalizeable and therefore hope they will have broad application for investigators using other FFQs who wish to add glycemic load values to their databases.

References

1. Jenkins DJ, Kendall CW, Augustin LS, Franceschi S, Hamidi M, Marchie A, Jenkins AL, Axelsen M. Glycemic index: Overview of implications in health and disease. *Am J Clin Nutr*. 2002;76(suppl):S266-S273.
2. Liu S, Willett WC, Stampfer MJ, Hu FB, Franz M, Sampson L, Hennekens CH, Manson JE. A prospective study of dietary glycemic load, carbohydrate intake, and risk of coronary heart disease in US women. *Am J Clin Nutr*. 2000;71:1455-1461.
3. Salmeron J, Manson JE, Stampfer MJ, Colditz GA, Wing AL, Willett WC. Dietary fiber, glycemic load, and risk of non-insulin dependent diabetes mellitus in women. *JAMA*. 1997;277:472-476.
4. Salmeron J, Ascherio A, Rimm EB, Colditz GA, Spiegelman D, Jenkins DJ, Stampfer MJ, Wing AL, Willett WC. Dietary fiber, glycemic load, and risk of NIDDM in men. *Diabetes Care*. 1997;20:545-550.
5. Terry PD, Jain M, Miller AB, Howe GR, Rohan TE. Glycemic load, carbohydrate intake, and risk of colorectal cancer in women: a prospective cohort study. *J Natl Cancer Inst*. 1995;95:914-916.
6. Jonas CR, McCullough ML, Teras LR, Walker-Thurmond KA, Thun MJ, Calle EE. Dietary glycemic index, glycemic load, and risk of incident breast cancer in postmenopausal women. *Cancer Epidemiol Biomarkers Prev*. 2003;12:573-577.
7. Higginbotham S, Zhang ZF, Lee IM, Cook NR, Buring JE, Liu S. Dietary glycemic load and breast cancer risk in the Women's Health Study. *Cancer Epidemiol Biomarkers Prev*. 2004;13:65-70.
8. Cho E, Spiegelman D, Hunter DJ, Chen WY, Colditz GA, Willett WC. Premenopausal dietary carbohydrate, glycemic index, glycemic load, and fiber in relation to risk of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2003;12:1153-1158.
9. Higginbotham S, Zhang ZF, Lee IM, Cook NR, Giovannucci E, Buring JE, Liu S. Dietary glycemic load and risk of colorectal cancer in the Women's Health Study. *J Natl Cancer Inst*. 2004;96:229-233.
10. Carnethon MR, Palaniappan LP, Burchfiel CM, Brancati FL, Fortmann SP. Serum insulin, obesity, and the incidence of type 2 diabetes in black and white adults: The atherosclerosis risk in communities study: 1987-1998. *Diabetes Care*. 2002;25:1358-1364.
11. Liu S, Manson JE, Stampfer MJ, Holmes MD, Hu FB, Hankinson SE, Willett WC. Dietary glycemic load assessed by food-frequency questionnaire in relation to plasma high-density-lipoprotein cholesterol and fasting plasma triacylglycerols in postmenopausal women. *Am J Clin Nutr*. 2001;73:560-566.
12. Buyken AE, Toeller M, Heitkamp G, Karamanos B, Rottiers R, Muggeo M, Fuller JH. Glycemic index in the diet of European outpatients with type 1 diabetes: Relations to glycated hemoglobin and serum lipids. *Am J Clin Nutr*. 2001;73:574-581.
13. Ford ES Liu S. Glycemic index and serum high-density lipoprotein cholesterol concentration among us adults. *Arch Intern Med*. 2001;161:572-576.
14. Frost G, Leeds AA, Dore CJ, Madeiros S, Brading S, Dornhorst A. Glycaemic index as a determinant of serum HDL-cholesterol concentration. *Lancet*. 1999; 353:1045-1048.
15. Liu S, Manson JE, Buring JE, Stampfer MJ, Willett WC, Ridker PM. Relation between a diet with a high glycemic load and plasma concentrations of high-sensitivity C-reactive protein in middle-aged women. *Am J Clin Nutr*. 2002;75:492-498.
16. Foster-Powell K, Holt SH, Brand-Miller JC. International table of glycemic index and glycemic load values. *Am J Clin Nutr*. 2002;76:5-56.
17. Subar AF, Midthune D, Kulldorff M, Brown CC, Thompson FE, Kipnis V, Schatzkin A. Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. *Am J Epidemiol*. 2000;152:279-286.
18. Pi-Sunyer FX. Glycemic index and disease. *Am J Clin Nutr*. 2002;76(suppl):S290-S298.
19. Rand WM, Pennington JAT, Murohy SP, Klensin JC. *Compiling Data for Food Composition Data Bases*. Tokyo, Japan: United Nations University Press; 1991.
20. Pennington JA, Schoen SA. Total diet study: Estimated dietary intakes of nutritional elements, 1982-1991. *Int J Vitam Nutr Res*. 1996;66:350-362.